

· 临床基础 ·

基于 Oncomine 和 TCGA 数据库分析 SLC2A1 基因在肺腺癌中的表达意义

李丹¹, 余涛², 曾智^{3*}, 吴杰¹, 叶鹏¹

1. 武汉大学人民医院 药学部, 湖北 武汉 430060

2. 华中科技大学同济医学院附属武汉中心医院 中西医结合肿瘤科, 湖北 武汉 4300142

3. 武汉大学人民医院 病理科, 湖北 武汉 430060

摘要: 目的 阐明 SLC2A1 基因在肺腺癌中的表达及临床意义。方法 通过检索 Oncomine 和 TCGA 等生物信息数据库中有关 SLC2A1 的信息, 并对所获取的数据资料挖掘并进行二次分析, 对 SLC2A1 在肺腺癌中的作用进行荟萃分析。结果 Oncomine 数据库中收集了 448 项不同类型 SLC2A1 的研究结果, 关于在肿瘤与对照组织中 SLC2A1 表达有统计学差异的结果有 47 项, 其中 SLC2A1 表达增高的有 43 项, 表达降低的有 4 项。肺腺癌中高表达的研究有 8 项、低表达的有 0 项。共有 8 项研究数据集涉及 SLC2A1 在肺腺癌组织和正常组织中的表达, 包括 786 例样本。在数据库中综合比较这 8 项研究成果, 发现与对照组相比, SLC2A1 在肺腺癌中的表达高于正常组织 ($P < 0.05$)。另外, 免疫组织化学显示 SLC2A1 在肺腺癌组织中表达较强或中等, 而在正常组织中表达较弱或呈阴性。TCGA 数据库中挖掘的结果也同样显示高表达 SLC2A1 的患者总体死亡率较高, 低表达 SLC2A1 的患者预后较好 ($P < 0.05$)。结论 基于公共数据库中肿瘤相关的基因信息, 提示 SLC2A1 的 mRNA 水平在肺腺癌组织中呈现高表达, 并与肺腺癌预后相关, 其有望成为肺腺癌药物治疗的重要治疗靶点。

关键词: SLC2A1 基因; 肺腺癌; Oncomine 数据库; TCGA 数据库

中图分类号: R979.1 文献标志码: A 文章编号: 1674-5515(2019)09-2807-06

DOI: 10.7501/j.issn.1674-5515.2019.09.052

Expression and clinical significance of SLC2A1 in lung adenocarcinoma: Analysis based on the data-mining of public database

LI Dan¹, YU Tao², ZENG Zhi³, WU Jie¹, YE Peng¹

1. Department of Pharmacy, Renmin Hospital of Wuhan University, Wuhan 430060, China

2. Department of Integrated Traditional Chinese and Western Medicine Ward of Oncology, the Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science & Technology, Wuhan 430014, China

3. Department of Pathology, Renmin Hospital of Wuhan University, Wuhan 430060, China

Abstract: Objective To explore the expression of SLC2A1 and its clinical significance in lung adenocarcinoma. **Methods** Data about SLC2A1 were retrieved from the Oncomine and TCGA database, then the role of SLC2A1 in lung adenocarcinoma was mined and analyzed. **Results** Totally 448 different types of cancer studies about the expression of SLC2A1 were identified in the Oncomine database. The SLC2A1 expression was statistically significant in 47 of the studies, over-expressed in 43 (8 in lung adenocarcinoma) and under-expressed in the other 4 (0 in lung adenocarcinoma). Eight datasets from eight different studies with 786 samples, addressed the expression of SLC2A1 in lung adenocarcinoma and normal tissues. The expression of SLC2A1 was significantly higher in the cancer than in the normal tissues ($P < 0.05$). In addition, immunohistochemistry showed that SLC2A1 was highly or moderately expressed in lung adenocarcinoma tissues, while it was weakly or negatively expressed in normal tissues. Furthermore, through mining and analyzing the data from TCGA database, the prognostic data showed a higher total mortality rate in the patients with higher

收稿日期: 2019-03-21

基金项目: 国家自然科学基金资助项目 (81803789、81704023、81602535); 湖北省卫计委中西医结合科研一般项目 (2017~2018-23); 武汉大学人民医院引导基金资助项目 (RMYD2018M79)

作者简介: 李丹, 副主任药师, 研究方向为中药抗肿瘤药理与临床药学。E-mail: lidandorren@126.com

*通信作者 曾智, 副主任医师, 研究方向为肿瘤的防治研究。E-mail: zhizeng@whu.edu.cn

SLC2A1 expression than in those with a lower expression ($P < 0.05$). **Conclusion** SLC2A1 is over-expressed in lung adenocarcinoma, and its expression is associated with the prognosis of lung adenocarcinoma, which may be used as an important target of medication for lung adenocarcinoma in the clinic.

Key words: SLC2A1 gene; lung adenocarcinoma; Oncomine database;TCGA database

肺癌因其高发病率、高死亡率以及病因复杂等原因,严重威胁着人类的健康与生命,一直以来都引起全球学者的广泛关注,一般肺癌可分为小细胞肺癌(SCLC,约占20%)和非小细胞肺癌(NSCLC,约占80%)^[1],肺腺癌是NSCLC最常见的组织学亚型,NSCLC 5年生存率低于15%,因此有必要继续研究其发生发展机制。

人类细胞膜上的葡萄糖转运载体(GLUTs)是介导葡萄糖顺浓度梯度进行跨膜转运的一类重要膜蛋白,对细胞能量供应和维持正常生命活动起到至关重要的作用。GLUT1由SLC2A1基因编码,位于1号染色体短臂区域。GLUT1已被证明在人类大多数癌症如脑癌、乳腺癌、头颈癌、膀胱癌、肾癌、结肠直肠癌、肺癌和卵巢癌等组织中呈现高表达^[2-3],对其研究较为深入。然而对其编码基因SLC2A1的研究则相对缺乏。近年来,多项研究显示SLC2A1的高表达可能与肺癌的发生、发展有关^[4-5],但由于研究方法、样本量和人群的差异,其结论可靠性需要更大样本进行验证。Oncomine和癌症基因组图谱(TCGA)数据库作为当前世界上最大的肿瘤基因芯片数据库和整合数据提取平台,其整合的文献及芯片数据以高质量得到学界的广泛认可,其最大的优势是丰富且规范的临床数据,以及针对每种癌型的大样本量,受到广泛科学研究工作者的信赖。本研究通过挖掘分析Oncomine和TCGA数据库中有关SLC2A1表达和肺腺癌患者临床病理特征及生存预后之间的关系,借此为肺腺癌的早期诊断和治疗提供更可信的科学依据。

1 资料与方法

1.1 从 Oncomine 和 TCGA 数据库提取数据

首先注册 Oncomine 数据库,获得使用数据库的权限,然后根据自己的需求在数据库中设定筛选和提取数据的条件。本研究中,在数据库检索界面输入重要的筛选条件:(1) cancer type: lung adenocarcinoma;(2) gene: SLC2A1;(3) data type: mRNA 或 DNA;(4) analysis type: cancer vs normal analysis。

接着,利用UALCAN平台进一步分析了TCGA数据库中SLC2A1在肺腺癌中的差异表达,且分析

了SLC2A1表达与患者临床病理特征(肿瘤分期、种族、年龄和性别)之间的关系。另外,在人类蛋白图谱项目中有关于正常组织、肿瘤和细胞系中蛋白表达谱的免疫组化图谱,使我们能够获得肺腺癌细胞水平SLC2A1表达模式的图谱。

为了评估SLC2A1在肺腺癌中的临床相关性,我们使用链接TCGA临床数据的UALCAN web工具。使用这个工具,绘制一个Kaplan-Meier图来探索SLC2A1水平与疾病预后的关系。

1.2 统计学分析

采用 t 检验来比较正常组织与肺腺癌病例组织之间SLC2A1表达的差异。采用Kaplan-Meier方法计算SLC2A1表达与肺腺癌预后的关系。所有统计学分析处理采用SPSS 18.0软件进行,作图采用Graphpad Prism 5.0软件进行,当 $P \leq 0.05$ 时认为有统计学显著性差异。

2 结果

2.1 SLC2A1 在不同类型肿瘤中的表达情况

Oncomine数据库中有关SLC2A1蛋白的不同类型研究结果共搜集了448项,涉及到肿瘤病例与对照组表达的比较、不同肿瘤类型之间的比较、肿瘤不同病理类型之间的比较等。关于SLC2A1表达有统计学差异的结果有47项研究,其中SLC2A1表达增高的有43项,表达降低的有4项。在肺癌中高表达的研究有8项、低表达的研究有0项。见表1。

2.2 SLC2A1 在肺腺癌中的表达

利用Oncomine数据库提取信息,2001年起共有8项研究,8个肺腺癌分型数据集涉及SLC2A1在肺腺癌和正常组织中的表达,共包括786例标本。相关研究分别发表于BMC Genomics、Genome Res、Plos One、Cancer Res、Nat Med、Plos One、Clin Cancer Res和Proc Natl Acad Sci USA。Oncomine数据库中综合比较这8个数据集的研究结果,对其进行二次分析发现,与对照组相比,SLC2A1在肺腺癌中呈现高表达($P < 0.05$),见图1。

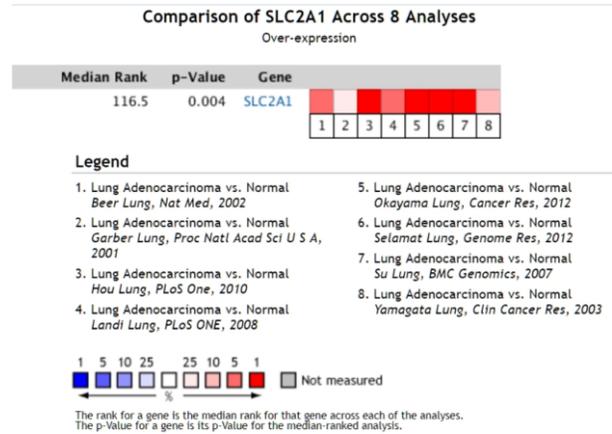
2.3 SLC2A1 在不同肺腺癌研究芯片中癌组织和正常组织之间的表达差异

Oncomine数据库中SLC2A1在不同肺腺癌研

表 1 SLC2A1 在 Oncomine 数据库肿瘤相关研究中的表达
Table 1 Expression of SLC2A1 in cancer-related studies in Oncomine database

肿瘤类型	表达上调的研究数目
膀胱癌	3
脑与中枢神经系统肿瘤	0
乳腺癌	10
宫颈癌	0
结肠癌	2
食管癌	2
胃癌	1
头颈癌	2
肾癌	5
白血病	2
肝癌	0
肺癌	8
淋巴瘤	1
黑色素瘤	0
骨髓瘤	0
其它肿瘤	2
卵巢癌	1
胰腺癌	4
前列腺癌	0
肉瘤	0
有显著性差异的研究总数	43

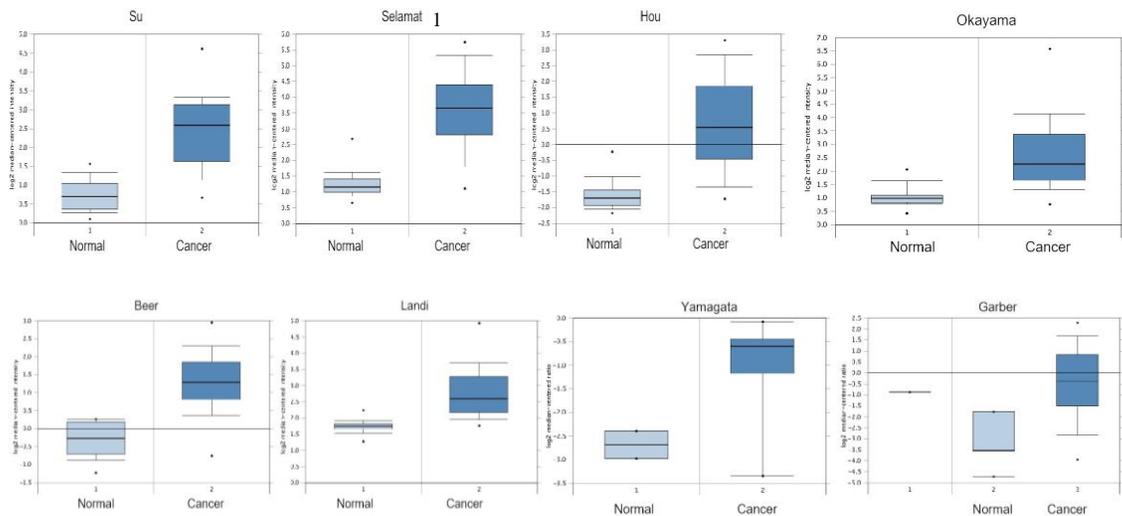
究芯片数据集中的肿瘤与对照组的表达结果见图 2。研究显示在 Su 等^[6]、Selamat 等^[7]、Hou 等^[8]、Okayama 等^[9]、Beer 等^[10]、Landi 等^[11]、Yamagata 等^[12]和 Garber 等^[13]这 8 个数据集研究中，SLC2A1 在肺腺癌中的表达量显著高均于正常组织 ($P < 0.05$)。不同研究中 8 个数据集对应的 SLC2A1 高表达倍数和各研究对应的 P 值见表 2。



1—8 分别表示 8 项研究结果，红色越深表示 SLC2A1 基因在该芯片中表达越高

1—8 represent the 8 studies on the expressions of SLC2A1 in lung adenocarcinoma, Darker red indicates higher SLC2A1 expression in the chips

图 1 SLC2A1 基因在 Oncomine 数据库中肺腺癌中的表达
Fig. 1 Expression of SLC2A1 gene in lung adenocarcinoma in the studies identified in the Oncomine database



与对照组比较: * $P < 0.05$
* $P < 0.05$ vs control group

图 2 Oncomine 数据库 SLC2A1 在不同肺腺癌研究数据集中的表达

Fig. 2 Expression of SLC2A1 m RNA in different studies identified in the Oncomine database

表 2 SLC2A1 基因在不同的研究者 8 个数据集集中的高表达情况
Table 2 Over-expression of SLC2A1 gene in the 8 datasets of different investigators

研究者	肿瘤类型	正常组织例数	癌症组织例数	P 值	改倍数变	发表杂志
Su	肺腺癌	30	27	1.12×10^{-10}	3.16	[6]
Selamat	肺腺癌	58	58	2.39×10^{-23}	5.15	[7]
Hou	肺腺癌	65	45	2.68×10^{-14}	4.80	[8]
Okayama	肺腺癌	20	226	5.86×10^{-21}	2.84	[9]
Beer	肺腺癌	10	86	5.99×10^{-8}	3.19	[10]
Landi	肺腺癌	49	58	2.15×10^{-16}	2.09	[11]
Yamagata	肺腺癌	2	9	0.008	3.16	[12]
Garber	肺腺癌	5	38	0.002	5.65	[13]

2.4 SLC2A1 蛋白在肺腺癌组织和正常组织之间的表达差异

利用 UALCAN 平台进一步分析了 TCGA 数据库中 SLC2A1 在肺腺癌中的差异表达情况，在 TCGA 数据库中涉及正常肺组织 59 例，肺腺癌组织 515 例，SLC2A1 基因在肺腺癌中的表达水平约为正常肺组织的 8.37 倍。为了通过临床病理分析来验证这一趋势，检测了人类蛋白图谱项目的免疫组化染色结果，发现 SLC2A1 在肺腺癌组织中表达较强或中等，而在正常组织中表达较弱或呈阴性。所有这些数据都证明 SLC2A1 的表达增加有助于肺腺癌的发生，表明其在肺腺癌中扮演促癌基因的作用。见图 3。

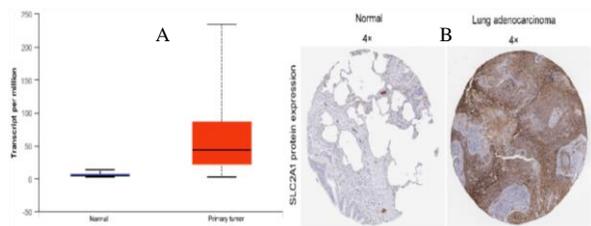


图 3 TCGA 数据库 (A) 和人类蛋白图谱项目 (B) 中 SLC2A1 在正常肺和肺腺癌组织中的表达情况

Fig. 3 SLC2A1 mRNA expression in normal tissues vs lung adenocarcinoma tissues from TCGA database (A) and human Project Atlas project (B)

2.5 SLC2A1 基因表达水平与临床病理特征之间的关系

利用 UALCAN 平台进一步分析了 TCGA 数据库中 SLC2A1 表达水平与患者临床病理特征之间的关系。研究结果表明 SLC2A1 表达水平与患者临床

病理特征如肿瘤分期、种族、年龄和性别没有相关性，见图 4。尽管男性患者 SLC2A1 表达高于女性患者 ($P < 0.01$)，但是差异为 1.28 倍，小于 1.5 倍，可以认为 SLC2A1 表达水平与性别之间不具有相关性。

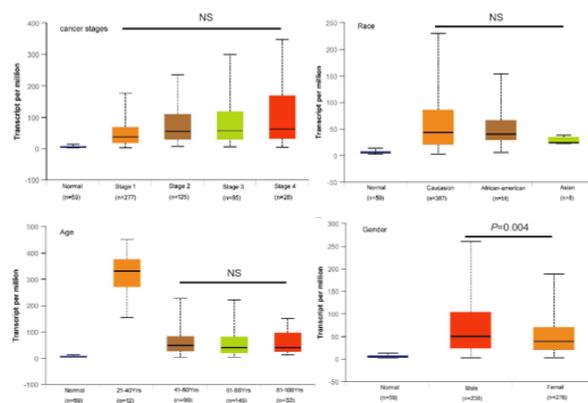


图 4 SLC2A1 基因表达与肺腺癌患者肿瘤分期 (A)、种族 (B)、年龄 (C)、性别 (D) 之间的关系

Fig. 4 Correlation between SLC2A1 mRNA expression and tumor stages (A), race (B), age (C), and gender (D)

2.6 SLC2A1 与肺腺癌预后之间的关系

通过上述数据库挖掘的信息发现 SLC2A1 在肺腺癌组织中表达升高，为了进一步明确 SLC2A1 表达与肺腺癌预后之间的关系，在 TCGA 数据库中挖掘提取 SLC2A1 基因的表达丰度值与肺腺癌预后相关的数据，利用 Kaplan-Meier 分析发现，SLC2A1 基因的表达水平与肺腺癌患者生存预后之间存在相关性，高表达 SLC2A1 的患者总体死亡率较高，低表达 SLC2A1 的患者预后较好 ($P < 0.0001$)，见图 5。

3 讨论

葡萄糖是动植物细胞所需要的基础代谢物质之

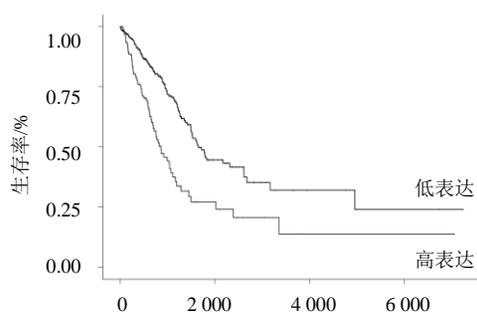


图5 SLC2A1 基因表达与肺腺癌患者预后之间的关系
Fig. 5 Relationship between the expression of SLC2A1 and the prognosis of lung adenocarcinoma

一，也是细胞主要代谢物质之间相互转化的中间产物和生物合成的原材料，肿瘤细胞也不例外^[14]。已有研究证实肿瘤细胞具有快速增殖的特点，代谢比正常细胞更加旺盛，需要大量物质供应和能量代谢作为支撑，因此需要摄取大量葡萄糖^[15]。然而葡萄糖是亲水性化合物，无法通过易化扩散的方式进入细胞内，需要特异性蛋白载体如 GLUT1 等介导其通过磷脂双分子层，从而转运至细胞内。因此肿瘤的恶性程度越高，增殖能力越强，摄取葡萄糖和糖酵解能力也越强，就会导致 GLUT1 的异常高表达，在肺腺癌中也是如此^[16]。本文所探讨的 SLC2A1 基因是 GLUT1 蛋白的编码基因，因此可以推测在肺腺癌中 SLC2A1 基因的表达也会被上调。

为证明该推测的准确性并进一步了解 SLC2A1 基因对肺腺癌预后的影响，本文采用了数据挖掘的方法，分析来自 Oncomine 和 TCGA 数据库中收录的不同研究者组成的大样本中 SLC2A1 基因的表达情况和相应患者的生存数据。众所周知，Oncomine 和 TCGA 数据库是目前世界上较大较全面的基因芯片数据库，其中收录保存的文献及芯片数据以其高质量和高可信度得到学界的广泛认可。本研究中我们通过收集并挖掘 Oncomine 数据库中收录的肺腺癌基因表达谱数据，总共纳入 8 个研究数据集，包含标本 786 例。结果发现 SLC2A1 基因在肺腺癌组织中的表达明显高于正常肺组织。又从 TCGA 数据库中筛选正常肺组织 59 例，肺腺癌组织 515 例，经统计分析得知 SLC2A1 基因在肺腺癌中的表达水平约为正常肺组织的 8.37 倍。进一步分析 SLC2A1 表达水平与患者临床病理特征的相关性发现，该基因表达和患者的肿瘤分期、种族、年龄和性别没有相

关性。Human Protein Atlas 数据库，简称 HPA 数据库，它提供了全部 24 000 种人类蛋白质的组织和细胞分布信息，这些受检组织来自 144 个不同个体和 216 个肿瘤组织，保证结果的准确性和充分的代表性。本文又通过检索 Human Protein Atlas 数据库了解 SLC2A1 基因编码蛋白的表达情况。结果发现 SLC2A1 在肺腺癌组织中表达较强或中等，而在正常组织中表达较弱或呈阴性，和基因表达分析结果一致。再通过 Kaplan-Meier 方法分析 TCGA 数据库中的生存数据发现，SLC2A1 基因的表达水平与肺腺癌患者生存预后之间存在相关性，其中高表达患者总体预后更差。

综上所述，高表达 SLC2A1 可作为肿瘤恶化进展和不良预后的指标，为临床诊疗提供参考，具备一定的临床应用价值。

参考文献

- [1] Janne P A, Yang J C, Kim D W, *et al.* AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer [J]. *N Engl J Med*, 2015, 372(18): 1689-1699.
- [2] Caillou B, Talbot M, Weyemi U, *et al.* Tumor-associated macrophages (TAMs) form an interconnected cellular supportive network in anaplastic thyroid carcinoma [J]. *PLoS One*, 2011, 6(7): e22567.
- [3] Hoskin P J, Sibtain A, Daley F M, *et al.* GLUT1 and CAIX as intrinsic markers of hypoxia in bladder cancer: relationship with vascularity and proliferation as predictors of outcome of ARCON [J]. *Br J Cancer*, 2003, 89(7): 1290-1297.
- [4] Yi J, Wei X, Li X, *et al.* A genome-wide comprehensive analysis of alterations in driver genes in non-small-cell lung cancer [J]. *Anticancer Drugs*, 2018, 29(1): 10-18.
- [5] Martinez-Terroba E, Behrens C, de Miguel F J, *et al.* A novel protein-based prognostic signature improves risk stratification to guide clinical management in early-stage lung adenocarcinoma patients [J]. *J Pathol*, 2018, 245(4): 421-432.
- [6] Su L J, Chang C W, Wu Y C, *et al.* Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme [J]. *BMC Genomics*, 2007, 8: 140.
- [7] Selamat S A, Chung B S, Girard L, *et al.* Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression [J]. *Genome Res*, 2012, 22(7): 1197-1211.
- [8] Hou J, Aerts J, den Hamer B, *et al.* Gene expression-

- based classification of non-small cell lung carcinomas and survival prediction [J]. *PLoS One*, 2010, 5(4): e10312.
- [9] Okayama H, Kohno T, Ishii Y, *et al.* Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas [J]. *Cancer Res*, 2012, 72(1): 100-111.
- [10] Beer D G, Kardias S L, Huang C C, *et al.* Gene-expression profiles predict survival of patients with lung adenocarcinoma [J]. *Nat Med*, 2002, 8(8): 816-824.
- [11] Landi M T, Dracheva T, Rotunno M, *et al.* Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival [J]. *PLoS One*, 2008, 3(2): e1651.
- [12] Yamagata N, Shyr Y, Yanagisawa K, *et al.* A training-testing approach to the molecular classification of resected non-small cell lung cancer [J]. *Clin Cancer Res*, 2003, 9(13): 4695-4704.
- [13] Garber M E, Troyanskaya O G, Schluens K, *et al.* Diversity of gene expression in adenocarcinoma of the lung [J]. *Proc Natl Acad Sci U S A*, 2001, 98(24): 13784-13789.
- [14] 高红亮, 丛威, 欧阳藩. 体外培养的哺乳动物细胞的葡萄糖和谷氨酰胺代谢 [J]. *生物技术通报*, 2000(2): 17-22.
- [15] 王红阳. 代谢异常与恶性肿瘤——也谈临床研究的科研思路 [J]. *医学研究生学报*, 2015, 28(7): 673-676.
- [16] Maki Y, Soh J, Ichimura K, *et al.* Impact of GLUT1 and Ki-67 expression on earlystage lung adenocarcinoma diagnosed according to a new international multidisciplinary classification [J]. *Oncol Rep*, 2013, 29(1): 133-140.